USPTO Serial Number: 10/014,744
Nong Ye *et al.*
Reply to Office Action dated July 16, 2004

Listing of the Claims:

1. (Currently amended) A method for classification of data, comprising:

providing first data from a physical process, the first data including a class label associated with attributes of the first data;

deriving a ~~dummy~~ temporary artificial cluster from centroid coordinates of the first data associated with the class label, wherein the temporary artificial ~~dummy~~ cluster has a class label different than the class label associated with the attributes of the first data;

determining distance measures between the first data and a plurality of clusters which include the temporary artificial ~~dummy~~ cluster;

creating a real cluster in the plurality of clusters if the first data is closest to the temporary artificial ~~dummy~~ cluster;

identifying a closest match between second data and the plurality of clusters; and

classifying the second data based on a class label of the closest match from the plurality of clusters.

2. (Previously presented) A method for classification of data, comprising:

providing first data from a physical process, the first data including a class label associated with attributes of the first data;

deriving a dummy cluster from centroid coordinates of the first data associated with the class label, wherein deriving the dummy cluster includes,

(a) creating first and second dummy clusters associated with first and second class labels, respectively~~+~~,

(b) initializing the first dummy cluster with the centroid coordinates of the first data if the class label of the first data matches the first class label~~+~~, and

(c) initializing the second dummy cluster with the centroid coordinates of the first data if the class label of the first data matches the second class label;

determining distance measures between the first data and a plurality of clusters which include the dummy cluster;

creating a real cluster in the plurality of clusters if the first data is closest to the dummy cluster;

identifying a closest match between second data and the plurality of clusters; and

classifying the second data based on a class label of the closest match from the plurality of clusters.

3. (Original) The method of claim 2, wherein deriving a dummy cluster further includes:

updating the centroid coordinates of the first dummy cluster for each first data having a class label that matches the first class label; and

updating the centroid coordinates of the second dummy cluster for each first data having a class label that matches the second class label.

4. (Previously presented) The method of claim 2, wherein determining distance measures between the first data and a plurality of clusters includes using one of the group of a

weighted Canberra distance, a weighted Euclidean distance, and a weighted Chi-squared distance for the distance measure.

5. (Previously presented) The method of claim 2, wherein determining distance measures between the first data and a plurality of clusters includes:

calculating sample variance of the first data;

calculating sample covariance of the first data;

calculating sample mean of the first data; and

calculating correlation coefficient from the sample variance, sample covariance, and sample mean of the first data.

6. (Previously presented) The method of claim 2, further including creating a real cluster in the plurality of clusters if the first data is closest to a cluster having a class label different than the class label associated with the first data.

7. (Previously presented) The method of claim 2, wherein identifying a closest match between second data and the plurality of clusters includes calculating a distance measure from one of the group of a weighted Canberra distance, a weighted Euclidean distance, and a weighted Chi-squared distance and using the closest distance measure as the closest match.

8. (Currently amended) A method of classifying first data from a physical process, comprising:

providing first data which includes a class label associated with attributes of the first data;

deriving ~~a dummy~~ an artificial cluster from centroid coordinates of the first data associated with the class label,

wherein the artificial ~~dummy~~ cluster has a class label different than the class label associated with the attributes of the first data to cause the first data to separate from the artificial cluster during the classification process;

determining distance measures between the first data and a plurality of clusters which include the artificial ~~dummy~~ cluster; and

creating a real cluster in the plurality of clusters if the first data is closest to the artificial ~~dummy~~ cluster.

9. (Original) The method of claim 8, further including:

identifying a closest match between second data and the plurality of clusters; and

classifying the second data based on a class label of the closest match from the plurality of clusters.

10. (Original) The method of claim 9, wherein identifying a closest match between second data and the plurality of clusters includes calculating a distance measure from one of the group of a weighted Canberra distance, a weighted Euclidean distance, and a weighted Chi-squared distance and using the closest distance measure as the closest match.

11. (Currently amended) The method of claim 8, wherein deriving the artificial ~~a dummy~~ cluster includes:

creating first and second artificial ~~dummy~~ clusters associated with first and second class labels, respectively;

initializing the first artificial ~~dummy~~ cluster with the centroid coordinates of the first data if the class label of the first data matches the first class label; and

initializing the second artificial ~~dummy~~ cluster with the centroid coordinates of the first data if the class label of the first data matches the second class label.

12. (Currently amended) The method of claim 11, wherein deriving the artificial ~~a dummy~~ cluster further includes:

updating the centroid coordinates of the first artificial ~~dummy~~ cluster for each first data having a class label that matches the first class label; and

updating the centroid coordinates of the second artificial ~~dummy~~ cluster for each first data having a class label that matches the second class label.

13. (Original) The method of claim 8, further including creating a real cluster in the plurality of clusters if the first data is closest to a cluster having a class label different than the class label associated with the first data.

14. (Previously presented) A method of classifying first data from a physical process, comprising:

providing first data which includes a class label associated with attributes of the first data;

deriving a dummy cluster from centroid coordinates of the first data associated with the class label, wherein the dummy cluster has a class label different than the class label associated with the attributes of the first data;

determining distance measures between the first data and a plurality of clusters which include the dummy cluster; and

creating a real cluster in the plurality of clusters if the first data is closest to a cluster having a class label different than the class label associated with the first data.

15. (Original) The method of claim 14 further including:

identifying a closest match between second data and the plurality of clusters; and

classifying the second data based on a class label of the closest match from the plurality of clusters.

16. (Original) The method of claim 15, wherein identifying a closest match between second data and the plurality of clusters includes calculating a distance measure from one of the group of a weighted Canberra distance, a weighted Euclidean distance, and a weighted Chi-squared distance and using the closest distance measure as the closest match.

17. (Original) The method of claim 14, wherein deriving a dummy cluster includes:

creating first and second dummy clusters associated with first and second class labels, respectively;

initializing the first dummy cluster with the centroid coordinates of the first data if the class label of the first data matches the first class label; and

initializing the second dummy cluster with the centroid coordinates of the first data if the class label of the first data matches the second class label.

18. (Original) The method of claim 17, wherein deriving a dummy cluster further includes:

updating the centroid coordinates of the first dummy cluster for each first data having a class label that matches the first class label; and

updating the centroid coordinates of the second dummy cluster for each first data having a class label that matches the second class label.

19. (Original) The method of claim 14, further including creating a real cluster in the plurality of clusters if the first data is closest to the dummy cluster.

20. (Currently amended) A digital storage medium encoded with a computer program which classifies data, the computer program comprising:

first instructions for providing first data from a physical process, the first data including a class label associated with attributes of the first data;

second instructions for deriving an artificial ~~a dummy~~ cluster from centroid coordinates of the first data associated with the class label, wherein the artificial ~~dummy~~ cluster has a class label different than the class label associated with the attributes of the first data to cause the first data to separate from the artificial cluster during the classification process;

third instructions for determining distance measures between the first data and a plurality of clusters which include the artificial ~~dummy~~ cluster; and

fourth instructions for creating a real cluster in the plurality of clusters if the first data is closest to the artificial ~~dummy~~ cluster.

21. (Original) The digital storage medium of claim 20 further including:

fifth instructions for identifying a closest match between second data and the plurality of clusters; and

sixth instructions for classifying the second data based on a class label of the closest match from the plurality of clusters.

22. (Currently amended) The digital storage medium of claim 20, wherein the second instructions further include:

fifth instructions for creating first and second artificial ~~dummy~~ clusters associated with first and second class labels, respectively;

sixth instructions for initializing the first artificial ~~dummy~~ cluster with the centroid coordinates of the first data if the class label of the first data matches the first class label; and

seventh instructions for initializing the second artificial ~~dummy~~ cluster with the centroid coordinates of the first data if the class label of the first data matches the second class label.

23. (Currently amended) The digital storage medium of claim 22, wherein the second instructions further include:

eighth instructions for updating the centroid coordinates of the first artificial ~~dummy~~ cluster for each first data having a class label that matches the first class label; and

ninth instructions for updating the centroid coordinates of the second artificial ~~dummy~~ cluster for each first data having a class label that matches the second class label.

24. (Original) The digital storage medium of claim 20, wherein the second instructions further include creating a real cluster in the plurality of clusters if the first data is closest to a cluster having a class label different than the class label associated with the first data.

25. (Previously presented) A method of classifying first data from a physical process, comprising:

providing first data which includes a class label associated with attributes of the first data;

forming a dummy cluster having a class label different than the class label associated with the attributes of the first data; and

creating a real cluster in a plurality of clusters which includes the dummy cluster if the first data is closest to a cluster having a class label different than the class label associated with the first data.

26. (Previously presented) The method of claim 25 further including:

identifying a closest match between second data and the plurality of clusters; and

classifying the second data based on a class label of the closest match from the plurality of clusters.

27. (Previously presented) The method of claim 26, wherein identifying a closest match between second data and the plurality of clusters includes calculating a distance measure from one of the group of a weighted Canberra distance, a

weighted Euclidean distance, and a weighted Chi-squared distance and using the closest distance measure as the closest match.

28. (Previously presented) The method of claim 25, wherein deriving a dummy cluster includes:

creating first and second dummy clusters associated with first and second class labels, respectively;

initializing the first dummy cluster with the centroid coordinates of the first data if the class label of the first data matches the first class label; and

initializing the second dummy cluster with the centroid coordinates of the first data if the class label of the first data matches the second class label.

29. (Currently amended) The method of claim 28 ~~29~~, wherein deriving a dummy cluster further includes:

updating the centroid coordinates of the first dummy cluster for each first data having a class label that matches the first class label; and

updating the centroid coordinates of the second dummy cluster for each first data having a class label that matches the second class label.

30. (Previously presented) The method of claim 25, further including creating a real cluster in the plurality of clusters if the first data is closest to the dummy cluster.